Simulation Theorems via Pseudo-random Properties

Arkadev Chattopadhyay [*] Michal Koucký [†] Bruno Loff [‡]
Sagnik Mukhopadhyay [§]

Abstract

We generalize the deterministic simulation theorem of Raz and McKenzie [RM99], to any gadget which satisfies certain hitting property. We prove that inner-product and gap-Hamming satisfy this property, and as a corollary we obtain deterministic simulation theorem for these gadgets, where the gadget's input-size is logarithmic in the input-size of the outer function. This yields the first deterministic simulation theorem with a logarithmic gadget size, answering an open question posed by Göös, Pitassi and Watson [GPW15].

Our result also implies the previous results for the Indexing gadget, with better parameters than was previously known. Moreover, logarithmic-sized gadget implies a quadratic separation in deterministic communication complexity and logarithm of partition number, no matter how high the partition number is with respect to the input size—something which is not achievable by previous results of [GPW15, AKK16].

Contents

[*] Tata Institute of Fundamental Research, Mumbai, `arkadev.c@tifr.res.in`
[†] Charles University, Prague, `koucky@iuuk.mff.cuni.cz`
[‡] INESC-TEC and University of Porto, Porto, `bruno.loff@gmail.com`
[§] Tata Institute of Fundamental Research, Mumbai, `sagnik@tifr.res.in`

1 Introduction

A very basic problem in computational complexity is to understand the *complexity* of a composed function $f \circ g$ in terms of the complexities of the two simpler functions f and g used for the composition. For concreteness, we consider $f : \{0,1\}^p \to \mathcal{Z}$ and $g : \{0,1\}^m \to \{0,1\}$ and denote the composed function as $f \circ g^p : \{0,1\}^{mp} \to \mathcal{Z}$; then f is called the *outer-function* and g is called the *inner-function.* The special case of $\mathcal{Z}$ being $\{0,1\}$ and f the XOR function has been the focus of several works [Yao82, Lev87, Imp95, Sha03, LSS08, VW08, She12b], commonly known as XOR lemmas. Another special case is when f is the trivial function that maps each point to itself. This case has also been widely studied in various parts of complexity theory under the names of 'direct sum' and 'direct product' problems, depending on the quality of the desired solution [JRS03, BPSW05, HJMR07, JKN08, Dru12, Pan12, JPY12, JY12, BBCR13, BRWY13a, BRWY13b, BBK$^+$13, BR14, KLL$^+$15, Jai15]. Making progress on even these special cases of the general problem in various models of computation is an outstanding open problem.

While no such general theorems are known, there has been some progress in the setting of communication complexity. In this setting the input for g is split between two parties, Alice and Bob. A particular instance of progress from a few years ago is the development of the pattern matrix method by Sherstov [She11] and the closely related block-composition method of Shi and Zhu [SZ09], which led to a series of interesting developments [Cha07, LSS08, CA08, She12a, She13, RY15], resolving several open problems along the way. In both these methods, the relevant analytic property of the outer function is approximate degree. While the pattern-matrix method entailed the use of a special inner function, the block-composition method, further developed by Chattopadhyay [Cha09], Lee and Zhang [LZ10] and Sherstov [She12a, She13], prescribed the inner function to have small discrepancy. These methods are able to lower bound the randomized communication complexity of $f \circ g^p$ essentially by the product of the approximate degree of f and the logarithm of the inverse of discrepancy of g.

The following simple protocol is suggestive: Alice and Bob try to solve f using a decision tree (randomized/deterministic) algorithm. Such an algorithm queries the input bits of f frugally. Whenever there is a query, Alice and Bob solve the relevant instance of g by using the best protocol for g. This allows them to progress with the decision tree computation of f, yielding (informally) an upper bound of $\mathcal{M}^{cc}\big(f \circ g^p\big) = O(\mathcal{M}^{dt}\big(f\big) \cdot \mathcal{M}^{cc}\big(g\big))$, where $\mathcal{M}$ could be the deterministic or randomized model and $\mathcal{M}^{dt}$ denotes the decision tree complexity [1]. A natural question is if the above upper bound is essentially optimal. The case when both f and g are XOR clearly shows that this is not always the case. However, this may just be a pathological case. Indeed it is natural to study for what models $\mathcal{M}$ and which inner functions g, is the above naive algorithm optimal.

In a remarkable and celebrated work, Raz and McKenzie [RM99] showed that this naïve upper bound is always optimal for *deterministic protocols*, when g is the Indexing function (IND), provided the *gadget size is polynomially large* in p. This theorem was the main technical workhorse of Raz and McKenzie to famously separate the monotone NC hierarchy. The work of Raz and McKenzie was recently simplified and built upon by Göös, Pitassi and Watson [GPW15] to solve a longstanding open problem in communication complexity. In line with [GPW15], we call such theorems *simulation theorems*, because they explicitly construct a decision-tree for f by simulating a given protocol for $f \circ g^p$. More recently, de Rezende, Nordström and Vinyals [dRNV16] port the above deterministic simulation theorem to the model of real communication, yielding new trade-offs for the measures of size and space in the cutting planes proof system.

We show a deterministic simulation theorem with improved gadget size by generalizing the simulation theorem of Raz-Mckenzie substantially. Our contribution in this part is two-fold.

[1] For randomized model, the upper bound holds with a multiplicative factor of $\log \mathcal{R}^q(f)$ — this is because we need to amplify the success probability of solving each instance of g so that we can do an union bound for the overall success probability of solving all instances of g.

On one hand, we generalize the proof considerably, by singling out a new pseudo-random property of a function $g : \{0,1\}^n \to \{0,1\}$, that we call "having (δ, h)-hitting rectangle-distributions", and then showing that a simulation theorem (i.e. , a theorem of the form of $\mathcal{D}^{cc}(f \circ g^p) = \Theta(\mathcal{D}^{dt}(f) \cdot h))$ will hold for any g with this property. Informally, a (δ, h)-distribution (for small δ and h) is a distribution over monochromatic rectangles such that a random rectangle from this distribution will intersect with any arbitrary large enough rectangle with good probability. By a function g having (δ, h)-hitting rectangle distribution, we mean that there are such 0 and 1-monochromatic rectangle distributions in g's truth-table. We then show that the inner-product function and the gap-Hamming problem have the above property. Mathematically, our main result is the following:

Theorem 1.1. If g has (δ, h)-hitting monochromatic rectangle-distributions, $\delta < 1/100$, and $p \leq 2^{\frac{h}{2}}$, then

$$\mathcal{D}^{dt}(f) \leq \frac{8}{h} \cdot \mathcal{D}^{cc}(f \circ g^p).$$

The techniques required to prove the deterministic simulation theorem are based on those that appear in [RM99, GPW15]. We show that two well-studied functions, — the inner-product function (IP) and the gap-Hamming family of functions (GH), — have the above property.

Theorem 1.2. Inner-product function and any function from the gap-Hamming class of promise-functions over n bits admit $(o(1), n/5)$-hitting monochromatic rectangle distributions.

Now, combining Theorem 1.1 and Theorem 1.2 immediately yields the following (*simulation*) theorem.

Theorem 1.3. Let $p \leq 2^{\frac{n}{200}}$, $f : \{0,1\}^p \to \mathcal{Z}$, where $\mathcal{Z}$ is any domain, and $g : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be inner-product function, or any function from the gap-Hamming class of promise-problems. Then,

$$\mathcal{D}^{cc}(f \circ g^p) = \Theta\bigg(\mathcal{D}^{dt}(f) \cdot n\bigg).$$

The inner-product function $\mathsf{IP}_n\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is defined as $\mathsf{IP}_n(x,y) = \sum_{i \in [n]} x_i \cdot y_i$, where the summation is taken over field $\mathbb{F}_2$. Problems in the class of gap-Hamming promise-problems, parameterized with γ and denoted by $\mathsf{GH}_{n,\gamma} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, distinguish the case of (x,y) having Hamming distance at least $(\frac{1}{2} + \gamma)n$ from the case of (x,y) having Hamming distance at most $(\frac{1}{2} - \gamma)n$, for $0 \leq \gamma \leq 1/4$. [2]

Note that this is the first deterministic simulation theorem with logarithmic gadget size, whereas the Raz-McKenzie simulation theorem requires a polynomial size gadget. This answers a problem raised by both Göös-Pittasi-Watson [GPW15] and Göös et.al. [GLM+15] of proving a Raz-McKenzie style deterministic simulation theorem for a different inner function than Indexing with a better gadget size. (Although the results presented in [GLM+15] do not deal with *deterministic* simulation theorems, the authors did raise the question of whether the proof of the deterministic simulation theorem can be simplified, and whether a simulation theorem can be shown for a lager class of gadgets g — we answer both these questions in this work.) Moreover, it is not hard to verify that an instance of the function g easily embeds in Indexing by exponentially blowing up the size. This enables us to also re-derive the original Raz-McKenzie simulation theorem for the Indexing function, even attaining significantly better parameters. This improvement in parameters answers a question posed to us recently by Jakob Nordström [Nor16]. In the next section, we will show how this strong form of simulation theorem helps us prove a strong complexity separation result.

It is well known that inner-product has strong pseudo-random properties. In particular it has vanishing discrepancy under the uniform distribution which makes it a good 2-source

[2] We remark here that the IND function also admits a $(\frac{1}{10}, \frac{3}{20}\log n)$-hitting monochromatic rectangle distribution—this is implicit in the proof of [GPW15, RM99]. Hence, Theorem 1.3 yields a simulation theorem for logarithmic sized IND gadget as well.

extractor. In fact, such strong properties of inner-product were recently used to prove simulation theorems for more exotic models of communication by Göös et al. [GLM$^+$15] and also by the authors and Dvořák [CDK$^+$17] to resolve a problem with a direct-sum flavor. By comparison, the pseudo-random property we abstract for proving our simulation theorem seems milder. This intuition is corroborated by the fact that we can show that gap-Hamming problems also possess our property, even though we know that these problems have large $\Omega(1)$ discrepancy under all distributions. Interestingly, any technique that relies on the inner-function having small discrepancy, such as the block-composition method, will not succeed in proving simulation theorems for such inner gadgets.

An application

The simulation theorem is used by [GPW15] to show a separation between the logarithm of partition number, χ, (as well as 1-partition number, χ_1) and deterministic communication complexity. At this point, it is interesting to note the relation between input size and the partition number of the functions for which they are able to show such a separation. For an input of size $N = p^{21}$ they exhibit a function that has has $\log(\chi_1)$ to be $\tilde{O}(\sqrt{p})$, whereas the deterministic communication complexity is $\tilde{\Omega}(p)$. They also exhibit another function for which $\log(\chi)$ is $\tilde{O}(p^{2/3})$ with the deterministic communication complexity being as high as $\tilde{\Omega}(p)$. It raises the question whether such a separation is possible when χ_1 (and, therefore, χ) is polynomially higher, i.e., say $\sqrt{N}$. Their results do not rule out the possibility that for all F such that $\log \chi_1(F)$ is $\omega(N^{\frac{1}{42}})$, the deterministic communication complexity of F is actually linear in partition complexity. Ambainis et al. [AKK16], in their improvement of [GPW15] in showing a near-optimal separation between $\log \chi$ and deterministic communication complexity, also leave this question open, as they use the simulation theorem of [GPW15] in a black-box fashion. Our result, with the improved gadget-size, rules out this possibility — our simulation theorem can be used (in the same way as in [GPW15] or [AKK16]) to show a function F^* for which $\log \chi_1(F^*)$ (or $\log \chi(F^*)$) is as big as $\tilde{O}(\sqrt{N})$ along with the deterministic communication complexity being at least square of that. More concretely, we show the following:

Theorem 1.4. For any function $s : \mathbb{Z} \to \mathbb{Z}$ such that $s(N) \leq \frac{\sqrt{N}}{\log N}$, there is a family of functions $\{F_N\}_{N\in\mathbb{Z}}$ such that $F_N : \{0,1\}^N \times \{0,1\}^N \to \mathcal{Z}$ has partition number $\chi(F_N) = 2^{\tilde{O}(s(N))}$ and deterministic communication complexity $\mathcal{D}^{cc}(F_N) \geq s(N)^2$.

We would, at this point, like to point out to the readers that a preliminary version of the results obtained in this paper appeared in [CKLM17].

We remark here that Wu, Yao and Yuen [WYY17] have independently reported a proof of the simulation theorem for the inner-product function, while a draft of this manuscript was already in circulation. Implicit in their proof is the construction of hitting rectangle-distributions for IP, and their construction of these distributions is similar to our own. This suggests that our pseudo-random property is essential to how all known deterministic simulation theorems are proven.

1.1 Our techniques

The main tool for proving a tight deterministic simulation theorem is to use the general framework of the Raz-McKenzie theorem as expounded by Göös-Pittasi-Watson [GPW15]. Given an input $z \in \{0,1\}^p$ for f, and wishing to compute $f(z)$, we will query the bits of z while simulating (in our head) the communication protocol for $f \circ g^p$, on inputs that are consistent with the queries to z we have made thus far. Namely, we maintain a rectangle $A \times B \subseteq \{0,1\}^{np} \times \{0,1\}^{np}$ so that for any $(x,y) \in A \times B$, $g^p(x,y)$ is *consistent* with z on all the coordinates that were queried. We will progress through the protocol with our rectangle

$A \times B$ from the root to a leaf. As the protocol progresses, $A \times B$ shrinks according to the protocol, and our goal is to maintain the consistency requirement. For that we need that inputs in $A \times B$ allow for all possible answers of g on those coordinates which we did not yet query. Hence $A \times B$ needs to be rich enough, and we are choosing a path through the protocol that affects this richness the least. If the protocol forces us to shrink the rectangle $A \times B$ so that we may not be able to maintain the richness condition, we query another coordinate of z to restore the richness. Once we reach a leaf of the protocol we learn a correct answer for $f(z)$, because there is an input $(x, y) \in A \times B$ on which $g^p(x, y) = z$ (since we preserved consistency) and all inputs in $A \times B$ give the same answer for $f \circ g^p$,

The technical property of $A \times B$ that we will maintain and which guarantees the necessary richness is called *thickness*. $A \times B$ is thick on the i-th coordinate if for each input pair $(x, y) \in A \times B$, even after one gets to see all the coordinates of x and y except for x_i and y_i, the *uncertainty* of what appears in the ith coordinate remains large enough so that $g(x_i, y_i)$ can be arbitrary. Let us denote by $\mathsf{Ext}^i_A(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$ the set of possible extensions x_i such that $\langle x_1, \dots, x_p \rangle \in A$. We define $\mathsf{Ext}^i_B(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_p)$ similarly. If for a given $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p$ and $y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_p$ we know that both $\mathsf{Ext}^i_A(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$ and $\mathsf{Ext}^i_B(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_p)$ are of size at least $2^{(\frac{1}{2}+\epsilon)n}$ then for $g = \mathsf{IP}_n$ there are extensions $x_i \in \mathsf{Ext}^i_A(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$ and $y_i \in \mathsf{Ext}^i_B(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_p)$ such that $\mathsf{IP}_n(x_i, y_i) = z_i$. Hence, we say that $A \times B$ is τ-thick if $\mathsf{Ext}^i_A(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$ and $\mathsf{Ext}^i_B(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_p)$ are of size at least $\tau \cdot 2^n$, for every choice of i and $x_1, \dots, x_p \in A$, $y_1, \dots, y_p \in B$.

So if we can maintain the thickness of $A \times B$, we maintain the necessary richness of $A \times B$. It turns out that this is indeed possible using the technique of Raz-McKenzie and Göös-Pittasi-Watson. Hence as we progress through the protocol we maintain $A \times B$ to be τ-thick and dense. Once the density of either A or B drops below certain level we are forced to make a query to another coordinate of z. Magically, that restores the density (and thus thickness) of $A \times B$ on coordinates not queried. (An intuitive reason is that if the density of extensions in some coordinate is low then the density in the remaining coordinates must be large.)

We give a sufficient condition for the inner function g that allows this type of argument to work, as follows. For $\delta \in (0, 1)$ and integer $h \geq 1$ we say that g has (δ, h)-*hitting monochromatic rectangle distributions* if there are two distributions σ_0 and σ_1 where for each $c \in \{0, 1\}$, σ_c is a distribution over c-monochromatic rectangles $U \times V \subset \{0, 1\}^n \times \{0, 1\}^n$ (i.e., $g(u, v) = c$ on every pair $(u, v) \in U \times V$), such that for any set $X \times Y \subset \{0, 1\}^n \times \{0, 1\}^n$ of sufficient size, a rectangle randomly chosen according to σ_c will intersect $X \times Y$ with large probability. More precisely, for any $c \in \{0, 1\}$ and for any $X \times Y$ with $|X|/2^n, |Y|/2^n \geq 2^{-h}$,

$$\Pr_{(U \times V) \sim \sigma_c} [(U \times V) \cap (X \times Y) \neq \varnothing] \geq 1 - \delta.$$

If such distributions σ_0 and σ_1 exist, we say that g has (δ, h)-hitting monochromatic rectangle-distributions.

The distribution σ_0 for $\mathsf{GH}_{n,\frac{1}{4}}$ is sampled as follows: we first sample a random string x of Hamming weight $\frac{n}{2}$, and we look at the set of all strings of Hamming weight $\frac{n}{2}$ which are at Hamming distance at most $\frac{n}{8}$ from x. Let's call this set U_x. The output of σ_0 will be the rectangle $U_x \times U_x$. The output of σ_1 is $U_x \times U_{\bar{x}}$, where $\bar{x}$ is the bit-wise complement of x. For any such x, $U_x \times U_x$ will be a 0-monochromatic rectangle and $U_x \times U_{\bar{x}}$ will be a 1-monochromatic rectangle. Note that if U_x does not hit a subset A of $\{0, 1\}^n$, then it means that x is at least $\frac{n}{8}$ Hamming distance away from the set A. By an application of Harper's theorem, we can show that for a sufficiently large set A, the number of strings which are at least $\frac{n}{8}$ Hamming distance away from A is exponentially small. This will imply that both σ_0 and σ_1 will hit a sufficiently large rectangle with probability exponentially close to 1, which is our required hitting property.

The σ_0 distribution for IP_n is picked as follows: To produce a rectangle $U \times V$ we sample uniformly at random a linear sub-space $V \subseteq F_2^n$ of dimension $n/2$ and we set $U = V^\perp$ to be the orthogonal complement of V. Since a random vector space of size $2^{n/2}$ hits a fixed subset of $\{0,1\}^n$ of size $2^{(\frac{1}{2}+\epsilon)n}$ with probability $1-O(2^{-\epsilon n})$, and both U and V are random vector spaces of that size, $U \times V$ intersects a given rectangle $X \times Y$ with probability $1-O(2^{-\epsilon n})$. Hence, we obtain $(O(2^{-\epsilon n}), (\frac{1}{2}+\epsilon)n)$-hitting distribution for IP. For the 1-monochromatic case, we first pick a random $a \in F_2^n$ of odd hamming weight and them pick random V and $U = V^{\perp}$ inside of the orthogonal complement of a. The distribution σ_1 outputs the 1-monochromatic rectangle $(a+V) \times (a+U)$, and will have the required hitting property.

1.2 Organization

Section 2 consists of basic definitions and preliminaries. In Section 3 we prove a deterministic simulation theorem for any gadget admitting (δ, h)-hitting monochromatic rectangle-distribution: sub-section 3.1 provides some supporting lemmas for the proof, and sub-section 3.2 holds the proof itself. In Section 4 we show that IND_n on n-bits has $(\frac{1}{10}, \frac{3}{20}\log n)$-hitting rectangle distribution, in Section 5 we show that $\mathsf{GH}_{n,\frac{1}{4}}$ on n-bits has $(o(1), \frac{n}{100})$-hitting rectangle distribution, and in Section 6 we show that IP on n-bits has $(o(1), n/5)$-hitting rectangle distribution.

2 Basic definitions and preliminaries

A *combinatorial rectangle*, or just a *rectangle* for short, is any product $A \times B$, where both A and B are finite sets. If $A' \subseteq A$ and $B' \subseteq B$, then $A' \times B'$ is called a *sub-rectangle* of $A \times B$. The *density* of A' in A is $\alpha = |A'|/|A|$.

Consider a product set $\mathcal{A} = \mathcal{A}_1 \times \ldots \times \mathcal{A}_p$, for some natural number $p \geq 1$, where each $\mathcal{A}_i$ is a subset of $\{0,1\}^n$. Let $A \subseteq \mathcal{A}$ and $I \subseteq [p] \stackrel{\text{def}}{=} \{1, \ldots, p\}$. Let $I = \{i_1 < i_2 < \cdot < i_k\}$, and $J = [p] \setminus I$. For any $a \in (\{0,1\}^n)^p$, we let $a_I = \langle a_{i_1}, a_{i_2}, \ldots, a_{i_k} \rangle$ be the projection of a onto the coordinates in I. Correspondingly, $A_I = \{a_I \mid a \in A\}$ is the projection of the entire set A onto I. For any $a' \in (\{0,1\}^n)^k$ and $a'' \in (\{0,1\}^n)^{p-k}$, we denote by $a' \times_I a''$ the p-tuple a such that $a_I = a'$ and $a_J = a''$. If $I = [k]$ for some $k \leq p$, we may omit the set I and write only $a' \times a''$. For $i \in [p]$ and a p-tuple a, $a_{\neq i}$ denotes $a_{[p]\setminus\{i\}}$, and similarly, $A_{\neq i}$ denotes $A_{[p]\setminus\{i\}}$. For $a' \in (\{0,1\}^n)^k$, we define the set of extensions $\mathsf{Ext}_A^J(a') = \{a'' \in (\{0,1\}^n)^{p-k} \mid a' \times_I a'' \in A\}$; we call those a'' *extensions* of a'. Again, if A and I are clear from the context, we may omit them and write only $\mathsf{Ext}(a')$.

Suppose $n \geq 1$ is an integer and $\mathcal{A} = \{0,1\}^n$. For an integer p, a set $A \subseteq \mathcal{A}^p$ and a subset $S \subseteq \mathcal{A}$, the restriction of A to S at coordinate i is the set $A^{i,S} = \{a \in A \mid a_i \in S\}$. We write $A_I^{i,S}$ for the set $(A^{i,S})_I$ (i.e. we first restrict the i-th coordinate then project onto the coordinates in I). Clearly $A_{\neq i}^{i,S}$ is non-empty if and only if S and A_i intersect.

The density of a set $A \subseteq \mathcal{A}^p$ will be denoted by $\alpha = \frac{|A|}{|\mathcal{A}|^p}$, and $\alpha_I^{i,S} = \frac{|A_I^{i,S}|}{|\mathcal{A}|^{|I|}}$.

Deterministic communication complexity

See [KN97] for an excellent exposition on this topic, which we cover here only very briefly. In the two-party communication model introduced by Yao [Yao79], two computationally unbounded players, Alice and Bob, are required to jointly compute a function $F : \mathcal{A} \times \mathcal{B} \to \mathcal{Z}$ where Alice is given $a \in \mathcal{A}$ and Bob is given $b \in \mathcal{B}$. To compute F, Alice and Bob communicate messages to each other, and they are charged for the total number of bits exchanged.

Formally, a *deterministic protocol* $\pi : \mathcal{A} \times \mathcal{B} \to \mathcal{Z}$ is a binary tree where each internal node v is associated with one of the players; Alice's nodes are labeled by a function $a_v : \mathcal{A} \to \{0,1\}$, and Bob's nodes by $b_v : \mathcal{B} \to \{0,1\}$. Each leaf node is labeled by an element of $\mathcal{Z}$. For each internal node v, the two outgoing edges are labeled by 0 and 1 respectively. The *execution* of π on the input $(a,b) \in \mathcal{A} \times \mathcal{B}$ follows a path in this tree: starting from the root, in each internal node v belonging to Alice, she communicates $a_v(a)$, which advances the execution to the corresponding child of v; Bob does likewise on his nodes, and once the path reaches a leaf node, this node's label is the output of the execution. We say that π *correctly computes* F on (a, b) if this label equals $F(a, b)$.

To each node v of a deterministic protocol π we associate a set $R_v \subseteq \mathcal{A} \times \mathcal{B}$ comprising those inputs (a, b) which cause π to reach node v. It is easy see that this set R_v is a combinatorial rectangle, i.e. $R_v = A_v \times B_v$ for some $A_v \subseteq \mathcal{A}$ and $B_v \subseteq \mathcal{B}$.

The *communication complexity of* π is the height of the tree. The *deterministic communication complexity of* F, denoted $\mathcal{D}^{cc}(F)$, is defined as the smallest communication complexity of any deterministic protocol which correctly computes F on every input.

Non-deterministic communication complexity

In the model of non-deterministic communication, we introduce another special player, namely the *prover*, who has access to both Alice and Bob's inputs, and whose job is to furnish a proof w (a binary string) to Alice and Bob witnessing $F(x, y) = 1$, whether or not it is the case. The *prover* is not trustworthy, and hence Alice and Bob has to verify the proof independently, and convince themselves whether it really is the case. If $F(x, y) = 0$, then at least one of the players should be able to detect that the proof is wrong. The *non-deterministic communication complexity of* F, denoted by $\mathcal{N}^{cc}(F)$, is defined to be the shortest length of the proof required for Alice and Bob to verify correctly on all inputs.

Covering. The measure of non-deterministic communication complexity can be viewed in the following combinatorial way: Consider the input space $\mathcal{A} \times \mathcal{B}$ (which is a combinatorial rectangle) and a covering of the inputs (x, y) for which $F(x, y) = 1$ by rectangles. These rectangles need not be disjoint, — the only conditions is that every (x, y) for which $F(x, y) = 1$ needs to lie in at least one rectangle, and no (x, y) for which $F(x, y) = 0$ lies in any of these rectangles. The *z-cover number*, denoted by $\mathsf{Cov}_z(F)$ where $z \in \{0, 1\}$, is the smallest number of rectangles required to cover the z-inputs of F. A little thinking shows that the non-deterministic communication complexity of F is exactly equal to the logarithm of 1-cover number of F, i.e., $\mathcal{N}^{cc}(F) = \log \mathsf{Cov}_1(F)$. The co-nondeterministic communication complexity of F, denoted by $co\mathcal{N}^{cc}(F)$, is, likewise, equal to $\log \mathsf{Cov}_0(F)$.

Partition number. As mentioned in the previous section, an interesting combinatorial property of the rectangle $\mathcal{A} \times \mathcal{B}$ is the *partition number* of the rectangle which is defined as follows:

Definition 2.1. The partition number of $F : \mathcal{A} \times \mathcal{B} \to \mathcal{Z}$ is defined by $\chi(F) = \sum_{z \in \mathcal{Z}} \chi_z(F)$, where $\chi_z(F)$ is the smallest number of rectangles needed to partition the inputs (x, y) such that $F(x, y) = z$ (i.e., the set $F^{-1}(z)$).

The quantity $\chi_z(F)$ is known as the *z-partition number of* F. The partition number of F is closely related to a variant of non-deterministic communication complexity of F. We call a non-deterministic protocol *unambiguous* if for each $(x, y) \in F^{-1}(1)$, there is exactly one proof that the *prover* can provide. The *unambiguous non-deterministic communication complexity* of F, denoted by $\mathcal{UN}^{cc}(F)$, is then equal to the logarithm of the 1-partition number of F, i.e., $\mathcal{UN}^{cc}(F) = \log \chi_1(F)$. Likewise, we can define $co\mathcal{UN}^{cc}(F) = \log \chi_0(F)$, and $2\mathcal{UN}^{cc}(F) = \log \chi(F)$, where $2\mathcal{UN}^{cc} = \mathcal{UN}^{cc} \cap co\mathcal{UN}^{cc}$.

Decision tree complexity

In the (Boolean) decision-tree model, we wish to compute a function $f : \{0, 1\}^p \to \mathcal{Z}$ when given query access to the input, and are charged for the total number of queries we make.

Formally, a *deterministic decision-tree* $T : \{0, 1\}^p \to \mathcal{Z}$ is a rooted binary tree where each internal node v is labeled with a variable-number $i \in [p]$, each edge is labeled 0 or 1, and and each leaf is labeled with an element of $\mathcal{Z}$. The execution of T on an input $z \in \{0,1\}^p$ traces a path in this tree: at each internal node v it queries the corresponding coordinate z_i, and follows the edge labeled z_i. Whenever the algorithm reaches a leaf, it outputs the associated label and terminates. We say that T *correctly computes* f on z if this label equals $f(z)$.

The *query complexity* of T is the height of the tree. The *deterministic query complexity* of f, denoted $\mathcal{D}^{dt}(F)$, is defined as the smallest query complexity of any deterministic decision-tree which correctly computes f on every input.

Functions of interest

The *Inner-product function on n-bits*, denoted IP_n is defined on $\{0,1\}^n \times \{0,1\}^n$ to be:

$$\mathsf{IP}_n(x,y) = \sum_{i\in[n]} x_i \cdot y_i \mod 2.$$

The *Indexing function on n-bits*, IND_n, is defined on $\{0,1\}^{\log n} \times \{0,1\}^n$ to be:

$$\mathsf{IND}_n(x,y) = y_x \quad \text{(the } x\text{'th bit of } y\text{)}.$$

Let n be a natural number and $\gamma = \frac{k}{n} \in (0,1/2)$. For two n-bit strings x and y, let $d_H(x,y) = \sum_i x_i \oplus y_i$ be their Hamming-distance. The *gap-Hamming problem*, denoted $\mathsf{GH}_{n,\gamma}$ is a promise-problem defined on $\{0,1\}^n \times \{0,1\}^n$, by the condition

$$\mathsf{GH}_{n,\gamma}(x,y) = \begin{cases} 1 & \text{if } d_H(x,y) \geq \left(\frac{1}{2}+\gamma\right)n, \\ 0 & \text{if } d_H(x,y) \leq \left(\frac{1}{2}-\gamma\right)n. \end{cases}$$

3 Deterministic simulation theorem

A *simulation theorem* shows how to construct a decision tree for a function f from a communication protocol for a composition problem $f \circ g^p$. Such a theorem can also be called a *lifting* theorem, if one wishes to emphasize that lower-bounds for the decision-tree complexity of f can be *lifted* to lower-bounds for the communication complexity of $f \circ g^p$. As mentioned in Section 1, the deterministic lifting theorem proved in [RM99], and subsequently simplified in [GPW15], uses IND_N as inner function g with N being polynomially larger than p. In this section we will show a deterministic simulation theorem for any function which possesses a certain pseudo-random property, which we will now define. Later we will show that the Inner-product and any function of gap-Hamming family have this property.

Definition 3.1 (Hitting rectangle-distributions)**.** Let $0 \leq \delta < 1$ be a real, $h \geq 1$ be an integer, and $\mathcal{A}, \mathcal{B}$ be some sets. A distribution σ over rectangles within $\mathcal{A} \times \mathcal{B}$ is called a (δ,h)*-hitting rectangle-distribution* if, for any rectangle $A \times B$ with $|A|/|\mathcal{A}|, |B|/|\mathcal{B}| \geq 2^{-h}$,

$$\Pr_{R\sim\sigma}[R \cap (A \times B) \neq \varnothing] \geq 1 - \delta.$$

Let $g : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ be a (possibly partial) function. A rectangle $A \times B$ is c-monochromatic with respect to g if $g(a,b) = c$ for every $(a,b) \in A \times B$.

Definition 3.2. For a real $\delta \geq 0$ and an integer $h \geq 1$, we say that a (possibly partial) function $g : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ *has* (δ,h)*-hitting monochromatic rectangle-distributions* if there are two (δ,h)-hitting rectangle-distributions σ_0 and σ_1, where each σ_c is a distribution over rectangles within $\mathcal{A} \times \mathcal{B}$ that are c-monochromatic with respect to g.

The theorem we will prove in Section 3.2 is the following:

Theorem 3.3 (Theorem 1.1 restated)**.** Let $\varepsilon \in (0,1)$ and $\delta \in (0, \frac{1}{100})$ be real numbers, and let $h \geq 6/\varepsilon$ and $1 \leq p \leq 2^{h(1-\varepsilon)}$ be integers. Let $f : \{0,1\}^p \to \mathcal{Z}$ be a function and $g : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ be a (possibly partial) function. If g has (δ, h)-hitting monochromatic rectangle-distributions then

$$\mathcal{D}^{dt}(f) \leq \frac{4}{\varepsilon \cdot h} \cdot \mathcal{D}^{cc}(f \circ g^p).$$

In Section 5 we will show that $\mathsf{GH}_{n,\frac{1}{4}}$ has $(o(1), \frac{n}{100})$-hitting monochromatic rectangle-distributions. From this we obtain a simulation theorem for $\mathsf{GH}_{n,\frac{1}{4}}$:

Corollary 3.4. Let n be large enough even integer, $\varepsilon \in (0,1)$, and $p \leq 2^{\frac{n}{100}(1-\varepsilon)}$ be an integer. For any function $f : \{0,1\}^p \to \{0,1\}$, $\mathcal{D}^{dt}(f) \leq \frac{400}{n\varepsilon} \cdot \mathcal{D}^{cc}(f \circ \mathsf{GH}^p_{n,\frac{1}{4}})$.

In Section 6 we will show that IP_n has $(o(1), n(\frac{1}{2} - \varepsilon))$-hitting monochromatic rectangle-distributions, for any constant $\varepsilon \in (0, 1/2)$. This allows us to derive after some simple calculations:

Corollary 3.5. Let n be large enough integer, $\varepsilon \in (0, 1/2)$ be a constant real, and $p \leq 2^{(\frac{1}{2}-\varepsilon)n}$ be an integer. For any function $f : \{0,1\}^p \to \{0,1\}$, $\mathcal{D}^{dt}(f) \leq \frac{36}{n\varepsilon} \cdot \mathcal{D}^{cc}(f \circ \mathsf{IP}^p_n)$.

These two corollaries together imply [3] Theorem 1.3. This allows us to significantly improve the gadget size known for simulation theorem of [RM99, GPW15], that uses the Indexing function instead of Inner-Product. Indeed, Jakob Nordström [Nor16] recently posed to us the challenge of proving a simulation theorem for $f \circ \mathsf{IND}^p_n$, with a gadget size n smaller than p^3 (p^3 is already a significant improvement to [RM99, GPW15]).

This follows from the above corollary, because of the following reduction: Given an instance $(a, b) \in (\{0,1\}^{mp})^2$ of $f \circ \mathsf{IP}^p_m$ where $p \leq 2^{m(\frac{1}{2}-\varepsilon)}$, Alice and Bob can construct an instance of $f \circ \mathsf{IND}^p_n$ where $n = 2^m$. Bob converts his input $b \in \{0,1\}^{mp}$ to $b' \in \{0,1\}^{np}$, so that each $b'_i = [\mathsf{IP}_n(x_1, b_i)\rangle, \cdots, \mathsf{IP}_n(x_n, b_i)\rangle]$ where $\{x_1, \cdots, x_n\} = \{0,1\}^m$ is an ordering of all m-bit strings. It is easy to see that $\mathsf{IP}_m(a_i, b_i) = \mathsf{IND}_n(a_i, b'_i)$. Hence it follows as a corollary to our result for IP:

Corollary 3.6. Let $\varepsilon \in (0, 1/2)$ be a constant real number, and n and p be sufficiently large natural numbers, such that $p \leq n^{\frac{1}{2}-\varepsilon}$. Then $\mathcal{D}^{dt}(f) = \frac{36}{\varepsilon \cdot \log n} \cdot \mathcal{D}^{cc}(f \circ \mathsf{IND}^p_n)$.

Also, it is worth noting that the proof of Lemma 7 in [GPW15], which Göös et al. call 'Projection Lemma', implicitly proves that IND_n has $(\frac{1}{150}, \frac{3}{20} \log n)$-hitting rectangle-distribution. Here the c-monochromatic rectangle distribution (c is either 1 or 0) is sampled as follows: Alice samples a subset of indices $U \subset [n]$ of size $n^{7/20}$, and Bob picks $V \subset \{0,1\}^n$ where $V = \{b \mid b_j = c \text{ for all } j \in U\}$.[4] Hence we can also apply Theorem 3.3 directly to obtain a corollary similar to Corollary 3.6 (albeit with much larger gadget size n). See Section 4 for a detailed derivation.

[3]The constant $\frac{1}{4}$ for $\mathsf{GH}^p_{n,\frac{1}{4}}$ in Corollary 3.4 is arbitrary. For any gap $\zeta \leq \frac{1}{2}$, we can show for $\mathsf{GH}^p_{n,\zeta}$ a $(2^{-n(1-H(\frac{1}{2}-\frac{\zeta}{4}))}, (1 - H(\frac{1}{2} - \frac{\zeta}{4}))n)$-hitting monochromatic distribution, where $H(\cdot)$ is the binary entropy function.

[4]Readers may note that δ in the proof of Claim 9 of [GPW15] is 1/4, where as we need $\delta < 1/100$. This is not a problem, as we can make δ as small a constant as we wish for by the same calculation as that in the proof of Claim 9.

3.1 Thickness and its properties

In this section, we list out a bunch of properties related to 'thickness' (a combinatorial property of a set that we will define below, — readers may also refer to [GPW15].) that we will need in Section 3.2 to prove a simulation theorem.

Definition 3.7 (Aux graph, average and min-degrees)**.** Let $p \geq 2$. For $i \in [p]$ and $A \subseteq \mathcal{A}^p$, the aux graph $G(A,i)$ is the bipartite graph with left side vertices A_i, right side vertices $A_{\neq i}$ and edges corresponding to the set A, i.e., (a', a'') is an edge iff $a' \times_{\{i\}} a'' \in A$.

We define the average degree of $G(A,i)$ to be the average right-degree:

$$d_{\text{avg}}(A,i) = \frac{|A|}{|A_{\neq i}|},$$

and the min-degree of $G(A,i)$, to be the minimum right-degree:

$$d_{\min}(A,i) = \min_{a' \in A_{\neq i}} |\mathsf{Ext}(a')|.$$

Definition 3.8 (Thickness and average-thickness)**.** For $p \geq 2$ and $\tau, \varphi \in (0,1)$, a set $A \subseteq \mathcal{A}^p$ is called τ*-thick* if $d_{\min}(A,i) \geq \tau \cdot |\mathcal{A}|$ for all $i \in [p]$. (Note, an empty set A is τ-thick.) Similarly, A is called φ*-average-thick* if $d_{\text{avg}}(A,i) \geq \varphi \cdot |\mathcal{A}|$ for all $i \in [p]$. For a rectangle $A \times B \subseteq \mathcal{A}^p \times \mathcal{B}^p$, we say that the rectangle $A \times B$ is τ-thick if both A and B are τ-thick. For $p = 1$, set $A \subseteq \mathcal{A}$ is τ-thick if $|A| \geq \tau \cdot |\mathcal{A}|$.

The following property is from [GPW15, Lemma 6].

Lemma 3.9 (Average-thickness implies thickness)**.** For any $p \geq 2$, if $A \subseteq \mathcal{A}^p$ is φ-average-thick, then for every $\delta \in (0,1)$ there is a $\frac{\delta}{p}\varphi$-thick subset $A' \subseteq A$ with $|A'| \geq (1-\delta)|A|$.

Proof. The set A' is obtained by running Algorithm 1.

Algorithm 1

1: Set $A^0 = A$, $j = 0$.
2: **while** $d_{\min}(A^j, i) < \frac{\delta}{p}\varphi \cdot 2^n$ for some $i \in [p]$ **do**
3: Let a' be a right node of $G(A^j, i)$ with non-zero degree less than $\frac{\delta}{p}\varphi \cdot 2^n$.
4: Set $A^{j+1} = A^j \setminus \{a'\} \times_i \mathsf{Ext}(a')$, i.e., remove every extension of a'. Increment j.
5: Set $A' = A^j$.

The total number of iteration of the algorithm is at most $\sum_{i \in [p]} |A_{\neq i}|$. (We remove at least one node in some $G(A^j, i)$ in each iteration which was a node also in the original $G(A,i)$.) So the number of iterations is at most

$$\sum_{i \in [p]} |A_{\neq i}| = \sum_{i \in [p]} \frac{|A|}{d_{\text{avg}}(A,i)} \leq \frac{p|A|}{\varphi 2^n}.$$

As the algorithm removes at most $\frac{\delta}{p}\varphi \cdot 2^n$ elements of A in each iteration, the total number of elements removed from A is at most $\delta|A|$, so $|A'| \geq (1-\delta)|A|$. Hence, the algorithm always terminates with a non-empty set A' that must be $\frac{\delta}{p}\varphi$-thick. $\square$

Lemma 3.10. Let $p \geq 2$ be an integer, $i \in [p]$, $A \subseteq \mathcal{A}^p$ be a τ-thick set, and $S \subseteq \mathcal{A}$. The set $A_{\neq i}^{i,S}$ is τ-thick. $A_{\neq i}^{i,S}$ is empty iff $S \cap A_i$ is empty.

Proof. Notice that $A^{i,S}_{\neq i}$ is non-empty iff $S \cap A_i$ is non-empty. Consider the case of $p \geq 3$. Let $a \in A$, where $a_i \in S$. Set $a' = a_{\neq i}$. For $j' \in [p-1]$, let $j = j'+1$ if $j' \geq i$, and $j = j'$ otherwise. Clearly, $\mathsf{Ext}^{\{j\}}_A(a_{\neq j}) \subseteq \mathsf{Ext}^{\{j'\}}_{A^{i,S}_{\neq i}}(a'_{\neq j'})$, hence the degree of a' in $G(A^{i,S}_{\neq i}, j')$ is at least the degree of a in $G(A, j)$ which is at least $\tau \cdot |\mathcal{A}|$. Hence, $A^{i,S}_{\neq i}$ is τ-thick.

To see the case $p = 2$, assume there is some string $a' \in A_{\neq i}$ which has some extension $a'' \in S$; but A itself is τ-thick, so there have to be at least $\tau \cdot |\mathcal{A}|$ many such a', which will then all be in $A^{i,S}_{\neq i}$. □

Lemma 3.11. Let $h \geq 1$, $p \geq 2$ and $i \in [p]$ be integers and $\delta, \tau, \varphi \in (0,1)$ be reals, where $\tau \geq 2^{-h}$. Consider a function $g : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ which has (δ, h)-hitting monochromatic rectangle-distributions. Suppose $A \times B \subseteq \mathcal{A}^p \times \mathcal{B}^p$ is a non-empty rectangle which is τ-thick, and suppose also that $d_{\mathrm{avg}}(A, i) \leq \varphi \cdot |\mathcal{A}|$. Then for any $c \in \{0,1\}$, there is a c-monochromatic rectangle $U \times V \subseteq \mathcal{A} \times \mathcal{B}$ such that

1. $A^{i,U}_{\neq i}$ and $B^{i,V}_{\neq i}$ is τ-thick,
2. $\alpha^{i,U}_{\neq i} \geq \frac{1}{\varphi}(1-3\delta)\alpha$,
3. $\beta^{i,V}_{\neq i} \geq (1-3\delta)\beta$,

where $\alpha = |A|/|\mathcal{A}|^p$, $\beta = |B|/|\mathcal{B}|^p$, $\alpha^{i,U}_{\neq i} = |A^{i,U}_{\neq i}|/|\mathcal{A}|^{p-1}$ and $\beta = |B^{i,U}_{\neq i}|/|\mathcal{B}|^{p-1}$.

The constant 3 in the statement may be replaced by any value greater than 2, so the lemma is still meaningful for δ arbitrarily close to 1/2.

Proof. Fix $c \in \{0,1\}$. Consider a matrix M where rows correspond to strings $a \in A_{\neq i}$, and columns correspond to rectangles $R = U \times V$ in the support of σ_c. Set each entry $M(a, R)$ to 1 if $U \cap \mathsf{Ext}^{\{i\}}_A(a) \neq \emptyset$, and set it to 0 otherwise.

For each $a \in A_{\neq i}$, $|\mathsf{Ext}^{\{i\}}_A(a)| \geq \tau|\mathcal{A}|$, and because σ_c is a (δ, h)-hitting rectangle-distribution and $\tau \geq 2^{-h}$, we know that if we pick a column R according to σ_c, then $M(a,R) = 1$ with probability $\geq 1-\delta$. So the probability that $M(a,R) = 1$ over uniform a and σ_c-chosen R is $\geq 1-\delta$.

Call a column of M A-*good* if $M(a,R) = 1$ for at least $1-3\delta$ fraction of the rows a. Now it must be the case that the A-good columns have strictly more than 1/2 of the σ_c-mass. Otherwise the probability that $M(a,R) = 1$ would be $< 1-\delta$.

A similar argument also holds for Bob's set $B_{\neq i}$. Hence, there is a c-monochromatic rectangle $R = U \times V$ whose column is both A-good and B-good in their respective matrices. This is our desired rectangle R.

We know: $|A^{i,V}_{\neq i}| \geq (1-3\delta)|A_{\neq i}|$ and $|B^{i,V}_{\neq i}| \geq (1-3\delta)|B_{\neq i}|$. Since $|B_{\neq i}| \geq |B|/|\mathcal{B}|$, we obtain $|B^{i,V}_{\neq i}|/|\mathcal{B}|^{p-1} \geq (1-3\delta)|B_{\neq i}|/|\mathcal{B}|^{p-1} \geq (1-3\delta)\beta$. Because $|A|/|A_{\neq i}| \leq \varphi|\mathcal{A}|$, we get

$$\frac{|A_{\neq i}|}{|\mathcal{A}|^{(p-1)}} \geq \frac{1}{\varphi} \cdot \frac{|A|}{|\mathcal{A}|^p} = \frac{\alpha}{\varphi}.$$

Combined with the lower bound on $|A^{i,V}_{\neq i}|$ we obtain $|A^{i,U}_{\neq i}|/|\mathcal{A}|^{p-1} \geq (1-3\delta)\alpha/\varphi$. The thickness of $A^{i,U}_{\neq i}$ and $B^{i,V}_{\neq i}$ follows from Lemma 3.10. □

Lemma 3.12. Let $p, h \geq 1$ be integers and $\delta, \tau \in (0,1)$ be reals, where $\tau \geq 2^{-h}$. Consider a function $g : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ which has (δ, h)-hitting monochromatic rectangle-distributions. Let $A \times B \subseteq \mathcal{A}^p \times \mathcal{B}^p$ be a τ-thick non-empty rectangle. Then for every $z \in \{0,1\}^p$ there is some $(a,b) \in A \times B$ with $g^p(a,b) = z$.

Proof. This follows from repeated use of Lemma 3.10. Fix arbitrary $z \in \{0,1\}^p$. Set $A^{(1)} = A$ and $B^{(1)} = B$. We proceed in rounds $i = 1, \ldots, p-1$ maintaining a τ-thick rectangle $A^{(i)} \times B^{(i)} \subseteq \mathcal{A}^{p-i+1} \times \mathcal{B}^{p-i+1}$. If we pick $U_i \times V_i$ from σ_{z_i}, then the rectangle

$(A^{(i)})_{\{i\}} \cap U_i \times (B^{(i)})_{\{i\}} \cap V_i$ will be non-empty with probability $\geq 1-\delta > 0$ (because σ_{z_i} is a (δ, h)-hitting rectangle-distribution and $\tau \geq 2^{-h}$). Fix such U_i and V_i. Set a_i to an arbitrary string in $(A^{(i)})_{\{i\}} \cap U_i$, and b_i to an arbitrary string in $(B^{(i)})_{\{i\}} \cap B_i$. Set $A^{(i+1)} = (A^{(i)})^{i,\{a_i\}}_{\neq i}$, $B^{(i+1)} = (B^{(i)})^{i,\{b_i\}}_{\neq i}$, and proceed for the next round. By Lemma 3.10, $A^{(i+1)} \times B^{(i+1)}$ is τ-thick.

Eventually, we are left with a rectangle $A^{(p)} \times B^{(p)} \subseteq \mathcal{A} \times \mathcal{B}$ where both $A^{(p)}$ and $B^{(p)}$ are τ-thick (and non-empty). Again with probability $1-\delta > 0$, the z_p-monochromatic rectangle $U_p \times V_p$ chosen from σ_{z_p} will intersect $A^{(p)} \times B^{(p)}$. We again set a_p and b_p to come from the intersection, and set $a = \langle a_1, a_2, \ldots, a_p\rangle$ and $b = \langle b_1, b_2, \ldots, b_p\rangle$. □

3.2 Proof of the simulation theorem

Now we are ready to present the simulation theorem (Theorem 3.3). Let $\varepsilon \in (0, 1/2)$ and $\delta \in (0, 1/100)$ be real numbers, and $h \geq 6/\varepsilon$ and $1 \leq p \leq 2^{h(1-\varepsilon)}$ be integers. Let $f : \{0,1\}^p \to \mathcal{Z}$ be a function and $g : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ be a (possibly partial) function. Assume that g has (δ, h)-hitting monochromatic rectangle-distributions. We assume we have a communication protocol Π for solving $f \circ g^p$, and we will use Π to construct a decision tree (procedure) for f. Let C be the communication cost of the protocol Π. If $p \leq 5C/h$ the theorem is true trivially. So assume $p > 5C/h$. Set $\varphi = 4 \cdot 2^{-\varepsilon h}$ and $\tau = 2^{-h}$. The decision-tree procedure is presented in Algorithm 2. On an input $z \in \{0,1\}^p$, it uses the protocol Π to decide which bits of z to query.

An informal description of simulation algorithm

Given an input $z \in \{0,1\}^p$, the algorithm starts traversing a path from the root of the protocol tree of Π. The variable v indicates the node of the protocol tree which is the current-node during the ongoing simulation. Associated with v, the algorithm maintains a rectangle $A \times B \subseteq \mathcal{A}^p \times \mathcal{B}^p$ and a set $I \subseteq [p]$ of indices. I corresponds to coordinates of the input z that were not queried, yet. Through out the execution of the algorithm, the following invariants are maintained: The set $A \times B$ is thick in the coordinates I, and every pair of inputs $(x, y) \in A \times B$ is consistent with the answer to the queries made so far. To start off, I is $[p]$, and $A \times B = \mathcal{A}^p \times \mathcal{B}^p$. So the invariants are trivially maintained at the beginning.

In each iteration of the simulation, the algorithm checks the following condition: Are both A_I and B_I φ-average-thick? Depending on the answer to this check, the algorithm does one of the following two things — If both A_I and B_I are φ-average-thick, the algorithm proceeds to that child of v which has at least half the mass of $A \times B$, and apply Lemma 3.9 to prune the rectangle associated with that child to ensure the thickness condition. Note that the working set $A \times B$ loses a constant fraction of density in doing so.

Otherwise, if there is a coordinate i in I, where A_I or B_I has low average degree, then the algorithm queries z_i and, depending on the value of z_i, applies Lemma 3.11 accordingly. Lemma 3.11 crucially exploits the fact that A_I and B_I is thick in i-th coordinate, and outputs a sub-rectangle of $A \times B$ which, in the i-th coordinate, is restricted to a z_i-monochromatic rectangle $U \times V$, while maintaining the thickness invariant in the coordinates $I \setminus \{i\}$. This also results in a boost in density of $A \times B$ in the current working universe $\mathcal{A}^{I \setminus i} \times \mathcal{B}^{I \setminus i}$. The algorithm updates I to be $I \setminus \{i\}$ and reiterates (i.e, does the average-thickness check again on $A \times B$ in the coordinate of the new I). We describe the parameters of the algorithm next in more detail.

Algorithm 2 Decision-tree procedure

Input: $z \in \{0,1\}^p$
Output: $f(z)$
1: Set v to be the root of the protocol tree for Π, $I = [p]$, $A = \mathcal{A}^p$ and $B = \mathcal{B}^p$.
2: **while** v is not a leaf **do**
3:     **if** A_I and B_I are both φ-average-thick **then**
4:         Let v_0, v_1 be the children of v.
5:         Choose $c \in \{0,1\}$ for which there is $A' \times B' \subseteq (A \times B) \cap R_{v_c}$ such that
6:           (1) $|A'_I \times B'_I| \geq \frac{1}{4}|A_I \times B_I|$
7:           (2) $A'_I \times B'_I$ is τ-thick. ▷ Using Lemma 3.9
8:         Update $A = A'$, $B = B'$ and $v = v_c$.
9:     **else if** $d_{\mathrm{avg}}(A_I, j) < \varphi|\mathcal{A}|$ for some $j \in [|I|]$ **then**
10:         Query z_i, where i is the j-th (smallest) element of I.
11:         Let $U \times V$ be a z_i-monochromatic rectangle of g such that
12:           (1) $A^{i,U}_{I\setminus\{i\}} \times B^{i,V}_{I\setminus\{i\}}$ is τ-thick,
13:           (2) $\alpha^{i,U}_{I\setminus\{i\}} \geq \frac{1}{\varphi}(1-3\delta)\alpha$,
14:           (3) $\beta^{i,V}_{I\setminus\{i\}} \geq (1-3\delta)\beta$, ▷ Using Lemma 3.11
15:         Update $A = A^{i,U}, B = B^{i,V}$ and $I = I \setminus \{i\}$.
16:     **else if** $d_{\mathrm{avg}}(B_I, j) < \varphi|\mathcal{B}|$ for some $j \in [|I|]$ **then**
17:         Query z_i, where i is the j-th (smallest) element of I.
18:         Let $U \times V$ be a z_i-monochromatic rectangle of g such that
19:           (1) $A^{i,U}_{I\setminus\{i\}} \times B^{i,V}_{I\setminus\{i\}}$ is τ-thick,
20:           (2) $\alpha^{i,U}_{I\setminus\{i\}} \geq (1-3\delta)\alpha$,
21:           (3) $\beta^{i,V}_{I\setminus\{i\}} \geq \frac{1}{\varphi}(1-3\delta)\beta$, ▷ Using Lemma 3.11
22:         Update $A = A^{i,U}, B = B^{i,V}$ and $I = I \setminus \{i\}$.
23: Output $f \circ g^p(A \times B)$.

Correctness. The algorithm maintains an invariant that $A_I \times B_I$ is τ-thick. This invariant is trivially true at the beginning.

If both A_I and B_I are φ-average-thick, the algorithm finds sets A' and B' on line 5–7 as follows. Consider the case that Alice communicates at node v. She is sending one bit. Let A_0 be inputs from A on which Alice sends 0 at node v and $A_1 = A \setminus A_0$. We can pick $c \in \{0,1\}$ such that $|(A_c)_I| \geq |A_I|/2$. Set $A'' = A_i$. Since A_I is φ-average-thick, A''_I is $\varphi/2$-average-thick. So using Lemma 3.9 on A''_I with δ set to $1/2$, we can find a subset A' of A'' such that A'_I is $\frac{\varphi}{4\cdot|I|}$-thick and $|A'_I| \geq |A''_I|/2$. ($A' \subseteq A''$ will be the pre-image of A'_I obtained from the lemma.) Since $\varphi = 4 \cdot 2^{-\varepsilon h}$ and $|I| \leq p \leq 2^{h(1-\varepsilon)}$, the set A'_I will be 2^{-h}-thick, i.e. τ-thick. Setting $B' = B$, the rectangle $A' \times B'$ satisfies properties from lines 6–7. A similar argument holds when Bob communicates at node v.

If A_I is not φ-average-thick, the existence of $U \times V$ at line 11 is guaranteed by Lemma 3.11. Similarly in the case when B_I is not φ-average-thick.

Next we argue that the number of queries made by Algorithm 2 is at most $5C/\varepsilon h$. In the first part of the **while** loop (line 3–8), the density of the current $A_I \times B_I$ drops by a factor 4 in each iteration. There are at most C such iterations, hence this density can drop by a factor of at most $4^{-C} = 2^{-2C}$. For each query that the algorithm makes, the density of the current $A_I \times B_I$ increases by a factor of at least $(1-3\delta)^2/\varphi \geq \frac{1}{2\varphi} \geq 2^{\varepsilon h - 3}$ (here we use the fact that

$\delta \le 1/100$). Since the density can be at most one, the number of queries is upper bounded by

$$\frac{2C}{\varepsilon h - 3} \le \frac{4C}{\varepsilon h}, \qquad \text{when } h \ge 6/\varepsilon.$$

Finally, we argue that $f(A \times B)$ at the termination of Algorithm 2 is the correct output. Given an input $z \in \{0,1\}^p$, whenever the algorithm queries any z_i, the algorithm makes sure that all the input pairs (x, y) in the rectangle $A \times B$ are such that $g(x_i, y_i) = z_i$ — because $U \times V$ is always a z_i-monochromatic rectangle of g. At the termination of the algorithm, I is the set of i such that z_i was not queried by the algorithm. As $p > 4C/\varepsilon h$, I is non-empty. Since $A_I \times B_I$ is τ-thick, it follows from Lemma 3.12 that $A \times B$ contains some input pair (x, y) such that $g^{|I|}(x_I, y_I) = z_I$, and so $g^p(x, y) = z$. Since Π is correct, it must follow that $f(z) = f \circ g^p(A \times B)$. This concludes the proof of correctness. $\square$

With greater care the same argument will allow for δ to be close to $\frac{1}{2}$. This would require also tightening the $1 - 3\delta$ factors appearing in Lemma 3.11 to something close to $1 - 2\delta$. The details are left to the readers, should they be interested.

4 Hitting rectangle-distribution for IND

Here we derive the $(\frac{1}{150}, \frac{3}{20}\log n)$-hitting monochromatic rectangle distribution for IND_n. Consider the following distribution σ_c over c-monochromatic rectangles: Alice samples a subset of indices $U \subset [n]$ of size $n^{7/20}$, and Bob picks $V \subset \{0,1\}^n$ where $V = \{b \mid b_j = c \text{ for all } j \in U\}$. We next show the following lemma.

Lemma 4.1. The distribution σ_c, for $c \in \{0,1\}$, is a $(\frac{1}{150}, \frac{3}{20}\log n)$-hitting c-monochromatic distribution for IND_n.

The proof of this lemma is implicit in the proof of Lemma 7 (Projection lemma) of [GPW15]. Göös et al. [GPW15] show [5] the following properties of σ_c in the course of proving their Lemma 7.

Lemma 4.2 ([GPW15])**.** If $U \times V$ is sampled from σ_c, then

1. For any set $A' \subseteq [n]$ that has size at least $n^{17/20}$, $\Pr_U[A' \cap U \neq \varnothing] \ge 1 - e^{-n^{1/5}}$,
2. For any set $B' \subseteq \{0,1\}^n$ with $\frac{|B'|}{2^n} \ge 2^{-n^{11/20}}$, $\Pr_U[V \cap B' \neq \varnothing] \ge \exp(-14(n^{-2/20} + n^{-6/20}))$.

The inverse exponential term on RHS is lower bounded by 3/4 in [GPW15]. We can bound this term by 199/200 as well. Hence, for this distribution, $\delta \le 1/200 + e^{-n^{1/5}} \le 1/150$.

Now we bound h. We have $\frac{|A'|}{n} \ge n^{-3/20} = 2^{-\frac{3}{20}\log n}$ from property (1). The bound on the size of B' comes from property (2), which is much smaller compared to $\frac{|A'|}{n}$. Hence we have $h = \frac{3}{20}\log n$.

5 Hitting rectangle-distributions for GH

We construct a hitting rectangle distribution for $\mathsf{GH}_{n,\frac{1}{4}}$. Subsequently, we will show a (δ, h)-hitting rectangle distribution where $\frac{|A \times B|}{|\{0,1\}^n \times \{0,1\}^n|} \ge 2^{-h}$.

Recall that $d_H(x, y)$ denotes the Hamming distance between the strings x and y. Let $B_r(x)$ be the Hamming ball of radius r around x, i.e. $B_r(x) = \{y \in \{0,1\}^n \mid d_H(x, y) \le r\}$; for a set $A \subset \{0,1\}^n$, $B_r(A) = \cup_{a \in A} B_r(a)$.

Let $\varepsilon = \frac{1}{8}$ and $\mathcal{H}$ be the set of all strings in $\{0,1\}^n$ with Hamming weight $n/2$. Now consider the rectangle distributions σ_0 and σ_1 obtained from the following sampling procedure:

[5] This is not stated as a separate lemma in [GPW15]. Property (1) is proven in the proof of property (0) of [GPW15], Property (2) is proven in the proof of Claim 9 of [GPW15]

- Choose a random string $x \in \mathcal{H}$, and let $\bar{x} \in \mathcal{H}$ be its bit-wise complement.
- Now let $U_x = B_{\varepsilon n}(x)$ and $V_x = B_{\varepsilon n}(\bar{x})$.
- The output of σ_1 is the rectangle $U_x \times V_x$, and the output of σ_0 is $U_x \times U_x$.

For the chosen value of ε, $U_x \times V_x$ is a 1-monochromatic rectangle, since for any $u \in U_x, v \in V_x$,

$$d_H(u,v) \geq n - 2\varepsilon n \geq \frac{3}{4}n.$$

On the other hand, $U_x \times U_x$ is 0-monochromatic, since for any $u, u' \in U_x$,

$$d_H(u,u') \leq 2\varepsilon n \leq \frac{1}{4}n.$$

Both inequalities are obtained by a straight-forward application of triangle inequality.

Lemma 5.1. The distributions σ_0 and σ_1 are $(2^{-\frac{n}{100}}, \frac{n}{100})$-hitting monochromatic rectangle distributions for $\mathsf{GH}_{n,\frac{1}{4}}$.

To prove Lemma 5.1, we need the following theorem due to Harper. We will call $S \subset \{0,1\}^n$ a *Hamming ball with center* $c \in \{0,1\}^n$ if $B_r(c) \subseteq S \subset B_{r+1}(c)$ for some non-negative integer r. For sets $S, T \subset \{0,1\}^n$, we define the *distance* between S and T as $d(S,T) = \min\{d_H(s,t) \mid s \in S, t \in T\}$.

Theorem 5.2 (Harper's theorem, [FF81, Har66])**.** Given any non-empty subsets S and T of $\{0,1\}^n$, there exist a Hamming ball S_0 with center $\bar{1}$ and Hamming ball T_0 with center $\bar{0}$ such that $|S| = |S_0|, |T| = |T_0|$ and $d(S_0, T_0) \geq d(S,T)$.

Note that Claim 5.2 also tells us when $B_r(S)$ is smallest for a set $S \subset \{0,1\}^n$. This can be argued in the following way: Given a set $S \in \{0,1\}^n$, let $T_S = \{0,1\}^n \setminus B_r(S)$. It is immediate that $d(S, T_S) = r+1$. Now let us suppose that S is such that it achieves the smallest $B_r(S')$ among all $S' \in \{0,1\}^n$ with $|S'| = |S|$. This also means that T_S is the biggest such set. Using Harper's theorem, we can find set S_0 and T_0 such that $d(S_0, T_0) \geq r+1$ where S_0 is centered around $\bar{1}$ and T_0 is centered around $\bar{0}$ with $|S_0| = |S|$ and $|T_0| = |T_S|$. Now it is easy to see that $T_0 \subseteq \{0,1\}^n \setminus B_r(S_0)$, i.e., $|T_S| = |T_0| \leq |T_{S_0}|$, which is a contradiction. This means that $|B_r(S)|$ will be the smallest if S is a Hamming ball centered around $\bar{1}$. This gives us the following corollary.

Corollary 5.3. For any non-negative integer $r \in [n]$ and among the set $\mathsf{A} = \{A \subset \{0,1\}^n \mid |A| = k\}$ for any k, if A is a Hamming ball centered around either $\bar{1}$ or $\bar{0}$, then $|B_r(A)| \leq |B_r(A')|$ for any $A' \in \mathsf{A}$.

Now we state the proof of Lemma 5.1.

Proof of Lemma 5.1. We will show that any set $A \subset \{0,1\}^n$ of size $|A| \geq 2^{\frac{99}{100}n}$ will be hit by U_x with probability $\geq 1 - 2^{-\frac{n}{100}}$. The lemma now follows since U_x and V_x have the same marginal distribution.

The event $U_x \cap A = \emptyset$ happens exactly when $x \notin B_{\varepsilon n}(A)$:

$$\Pr_x[U_x \cap A = \emptyset] = \Pr_x[x \notin B_{\varepsilon n}(A)] \leq \frac{2^n - |B_{\varepsilon n}(A)|}{2^n}.$$

From Corollary 5.3 we know that $|B_{\varepsilon n}(A)|$ is smallest when A is itself a Hamming ball around 0 of the same density as A. I.e., if $|B_{\gamma n}(0)| \leq |A|$, then

$$|B_{\varepsilon n}(A)| \geq |B_{\varepsilon n}(B_{\gamma n}(0))| = |B_{(\gamma+\varepsilon)n}(0)|.$$

For $\gamma = \frac{1}{2} - \frac{\varepsilon}{2} = \frac{1}{2} - \frac{1}{16}$, and since $H(\gamma) < \frac{99}{100}$, we have

$$|B_{\gamma n}(0)| \leq 2^{H(\gamma)n} \leq 2^{\frac{99}{100}n} \leq |A|.$$

And so $|B_{\varepsilon n}(A)| \geq |B_{(\gamma+\varepsilon)n}(0)| = |B_{\frac{n}{2}+\frac{n}{16}}(0)| \geq 2^n - |B_{\frac{n}{2}-\frac{n}{16}}(1)| \geq 2^n - 2^{\frac{99}{100}n}$. It now follows

$$\Pr_x[U_x \cap A = \emptyset] \leq \frac{2^{\frac{99}{100}n}}{2^n} \leq 2^{-\frac{n}{100}}. \qquad \square$$

6 Hitting rectangle-distributions for IP

In this section, we will show that IP_n has $(4 \cdot 2^{-n/20}, n/5)$-hitting monochromatic rectangle-distributions. This will show a deterministic simulation result when the inner function is IP_n, i.e.,

$$\mathcal{D}^{cc}(f \circ \mathsf{IP}_n^p) \geq \mathcal{D}^{dt}(f) \cdot \Omega(n).$$

We will use the following well-known variant of Chebyshev's inequality:

Proposition 6.1 (Second moment method)**.** Suppose that $X_i \in [0,1]$ and $X = \sum_i X_i$ are random variables. Suppose also that for all i and j, X_i and X_j are *anti-correlated*, in the sense that

$$\mathbf{E}[X_i X_j] \leq \mathbf{E}[X_i] \cdot \mathbf{E}[X_j].$$

Then X is well-concentrated around its mean, namely, for every ε:

$$\Pr[X \in \mu(1 \pm \varepsilon)] \geq 1 - \frac{1}{\varepsilon^2 \mu}.$$

All of the rectangle-distributions rely on the following fundamental anti-correlation property:

Lemma 6.2 (Hitting probabilities of random subspaces)**.** Let $0 \leq d \leq n$ be natural numbers. Fix any $v \neq w$ in $\mathbb{F}_2^n$, and pick a random subspace V of dimension d. Then the probability that $v \in V$ is exactly

$$p_v = \begin{cases} \frac{2^d-1}{2^n-1} & \text{if } v \neq 0 \\ 1 & \text{if } v = 0. \end{cases}$$

And the probability that both $v, w \in V$ is exactly

$$p_{v,w} = \begin{cases} \binom{2^d-1}{2} \Big/ \binom{2^n-1}{2} & \text{if } v, w \neq 0 \\ p_v & \text{if } w = 0, \text{ and} \\ p_w & \text{if } v = 0. \end{cases}$$

Hence it always holds that $p_{v,w} \leq p_v p_w$.

Proof. The case when v or w are 0 is trivial. The value $p_v = \Pr[v \in V]$ for a random subspace V of dimension d equals $\Pr[Mv = 0]$ for a random non-singular $(n-d) \times n$ matrix M, letting $V = \ker M$. For any $v \neq 0, v' \neq 0$, M will have the same distribution as MN, where N is some fixed linear bijection of F_2^n mapping v to v'; it then follows that $p_v = p_{v'}$ always. But then

$$\sum_{v \neq 0} p_v = \mathbf{E}\left[\sum_{v \neq 0} [v \in V]\right] = 2^d - 1,$$

and since all p_v's are equal, then $p_v = \frac{2^d-1}{2^n-1}$.

Now let $p_{v,w} = \Pr[v \in V, w \in V]$. In the same way we can show that $p_{v,w} = p_{v',w'}$ for all two such pairs, since a linear bijection will exist mapping v to v' and w to w' (because every $v \neq w$ is linearly independent in $\mathbb{F}_2^n$). And now

$$\sum_{v,w \neq 0} p_{v,w} = \mathbf{E}\left[\sum_{v,w \neq 0} [v \in V][w \in V]\right] = \binom{2^d - 1}{2}.$$

The value of $p_{v,w}$ is then as claimed. We conclude by estimating

$$\frac{p_{v,w}}{p_v p_w} = \frac{\binom{2^d-1}{2}}{\binom{2^n-1}{2}} \cdot \frac{1}{p_v p_w} = \frac{2^d - 2}{2^d - 1} \cdot \frac{2^n - 1}{2^n - 2} < 1. \quad \square$$

It can now be shown that a random subspace of high dimension will hit a large set w.h.p.:

Lemma 6.3. Let $\varepsilon < \frac{1}{2}$ be a positive real number, and consider a set $B \subseteq \{0,1\}^n$ of density $\beta = \frac{|B|}{2^n} \geq 2^{-(\frac{1}{2}-\varepsilon)n}$. Pick V to be a random linear subspace of $\{0,1\}^n$ of dimension d, where $d \geq (\frac{1}{2} - \frac{\varepsilon}{4})n + 6$. Then

$$\Pr_V\left[\frac{|B \cap V|}{|V|} \in (1 \pm 2^{-\frac{\varepsilon}{4}n}) \cdot \beta\right] \geq 1 - \frac{1}{4} \cdot 2^{-\frac{\varepsilon}{4}n}.$$

Proof. Let $b_1, \ldots, b_N$ be the elements of B, and define the random variables $X_i = [b_i \in V]$ and $X = |B \cap V| = \sum_i X_i$. The $\mathbf{E}[X_i]$ were computed in the proof of Lemma 6.2, which gives us

$$\mu = \mathbf{E}[X] = \sum_i \mathbf{E}[X_i] = \begin{cases} \beta 2^n \frac{2^d-1}{2^n-1} & \text{if } \bar{0} \notin B \\ \beta 2^n \frac{2^d-1}{2^n-1} + (1 - \frac{2^d-1}{2^n-1}) & \text{otherwise.} \end{cases}$$

Let's look at the case where $\bar{0} \notin B$. We can estimate μ as follows:[6]

$$\mu \;=\; \left(1 + \frac{1}{2^n - 1}\right)(1 - 2^{-d})\beta|V| \;\in\; (1 \pm 2^{-(\frac{1}{2}-\frac{\varepsilon}{2})n})^2 \beta|V| \;\subseteq\; \left(1 \pm \frac{1}{3} \cdot 2^{-\frac{\varepsilon}{2}n}\right)\beta|V|.$$

When $\bar{0} \in B$ we still have $\mu \in (1 \pm 2^{-\frac{\varepsilon}{2}n})\beta|V|$, because $1 - \frac{2^d-1}{2^n-1} \leq 1 \ll \frac{1}{3} \cdot 2^{-\frac{\varepsilon}{2}n}\beta|V|$. So this holds in both cases.

Lemma 6.2 also says that $\mathbf{E}[X_i X_j] \leq \mathbf{E}[X_i]\mathbf{E}[X_j]$ for all $i \neq j$. And so by the second moment method (Lemma 6.1):

$$\Pr\left[X \in \mu\left(1 \pm \delta\right)\right] \geq 1 - \frac{1}{\delta^2 \mu}$$

which means,

$$\Pr\left[X \in (1 \pm 2^{-\frac{\varepsilon}{2}n})(1 \pm \delta)\beta|V|\right] \geq 1 - \frac{1}{\delta^2 \cdot \beta \cdot 2^d \cdot (1 - 2^{-\frac{\varepsilon}{2}n})}$$

Taking $\delta = \frac{1}{3}2^{-\frac{\varepsilon}{4}n}$, we get,

$$\Pr\left[X \in (1 \pm 2^{-\frac{\varepsilon}{4}n})\beta|V|\right] \geq 1 - \frac{9}{2^{-\frac{\varepsilon}{2}n} \cdot 2^{-(\frac{1}{2}-\varepsilon)n} \cdot 64 \cdot 2^{(\frac{1}{2}-\frac{\varepsilon}{4})n}} \geq 1 - \frac{1}{4} \cdot 2^{-\frac{\varepsilon}{4}n}. \;\square$$

We will show a similar result when we pick the set V in the following manner: First we pick a uniformly random odd-Hamming weight vector $a \in \{0,1\}^n$, and then we pick W to be a random subspace of dimension d within $a^\perp$, where $d \geq (\frac{1}{2} - \frac{\varepsilon}{4})n + 6$; then $V = a + W$.

[6] Throughout the proof we will use the fact that $(1 \pm \delta)^2 \subseteq 1 \pm 3 \cdot \delta$, and also that $1 \pm \delta \subseteq 1 \pm \delta'$ whenever $\delta \leq \delta'$.

Lemma 6.4. Consider a set $B \subseteq \{0,1\}^n$ of density $\beta = \frac{|B|}{2^n} \geq 2^{-(\frac{1}{2}-\varepsilon)n}$. Pick V as described above. Then

$$\Pr_V\left[\frac{|B \cap V|}{|V|} \in \beta(1 \pm 2^{-\frac{\varepsilon}{4}n})\right] \geq 1 - 2^{-\frac{\varepsilon}{4}n}.$$

Proof. Let $B' = (B - a) \cap a^\perp$ and let $\beta' = \frac{|B'|}{|a^\perp|}$. A string $a \in \{0,1\}^n$ is called *good* when

$$\beta' \stackrel{\text{def}}{=} \frac{|(B-a) \cap a^\perp|}{|a^\perp|} \in \beta \cdot (1 \pm 2^{-\frac{\varepsilon}{4}n}).$$

We will later show that if a is a uniformly random string of odd Hamming weight, then

$$\Pr_a[a \text{ is good}] \geq 1 - \frac{2}{4} \cdot 2^{-\frac{\varepsilon}{4}n}. \tag{$*$}$$

For every good a, Lemma 6.3 gives us:

$$\Pr_W\left[\frac{|B' \cap W|}{|W|} \in \beta'(1 \pm 2^{-\frac{\varepsilon}{4}n}) \,\middle|\, a\right] \geq 1 - \frac{1}{4} \cdot 2^{-\frac{\varepsilon}{4}n}.$$

Our result then follows by Bayes' rule.

To prove ($*$), suppose that a is chosen to be a uniformly random non-zero string (i.e. with either even or odd Hamming weight). Then $a^\perp$ is a uniformly random subspace of dimension $n - 1 \gg (\frac{1}{2} - \frac{\varepsilon}{4})n + 6$. Hence by Lemma 6.3,

$$\Pr_a\left[\frac{|B \cap a^\perp|}{|a^\perp|} \in \beta \cdot (1 \pm 2^{-\frac{\varepsilon}{4}n})\right] \geq 1 - \frac{1}{4} \cdot 2^{-\frac{\varepsilon}{4}n}. \tag{$**$}$$

Now $|a^\perp| = 2^{n-1}$, so if $a^\parallel$ denotes the complement of $a^\perp$ (in $\{0,1\}^n$), then $|a^\parallel| = 2^{n-1}$ also, and

$$\frac{|B \cap a^\perp|}{|a^\perp|} \in \beta \cdot (1 \pm 2^{-\frac{\varepsilon}{4}n}) \iff |B \cap a^\perp| \in \frac{1}{2}|B| \cdot (1 \pm 2^{-\frac{\varepsilon}{4}n}) \iff \frac{|B \cap a^\parallel|}{|a^\parallel|} \in \beta \cdot (1 \pm 2^{-\frac{\varepsilon}{4}n}).$$

So ($**$) also holds with respect to the rightmost (equivalent) event. Since a uniformly random non-zero a has odd Hamming weight with probability $> \frac{1}{2}$, it must then follow that if we pick a uniformly random a with odd Hamming weight, then:

$$\Pr_a\left[\frac{|B \cap a^\parallel|}{|a^\parallel|} \in \beta \cdot (1 \pm 2^{-n/20})\right] \geq 1 - \frac{2}{4} \cdot 2^{-\frac{\varepsilon}{4}n}.$$

Now notice that $|a^\parallel| = |a^\perp|$ and that for odd Hamming weight a, $B \cap a^\parallel = (B - a) \cap a^\perp$; this establishes ($*$). $\square$

The lemmas above are the key to constructing rectangle-distributions for IP.

Lemma 6.5. For all $0 < \varepsilon < 1/2$ and every sufficiently large n, IP_n has $(2 \cdot 2^{-\frac{\varepsilon}{4}n}, (\frac{1}{2} - \varepsilon)n)$-hitting monochromatic rectangle-distributions.

Proof. We define the distributions σ_0 and σ_1 by the following sampling methods:

Sampling from σ_0: We choose a uniformly-random $\frac{n}{2}$-dimensional subspaces V of $\mathbb{F}_2^n$, and let $V^\perp$ be its orthogonal complement; output $V \times V^\perp$.

Sampling from σ_1: First we pick $a \in \{0,1\}^n$ uniformly at random conditioned on the fact that a has odd Hamming weight; then we pick random subspace W of dimension $(n-1)/2$ from $a^\perp$, and let $W^\perp$ be the orthogonal complement of W *inside* $a^\perp$. We output $V \times V^\parallel$, where $V = a + W$ and $V^\parallel = a + W^\perp$.

The rectangles produced above are monochromatic as required. Also, V and $V^\perp$ of σ_0 are both random subspaces of dimension $\geq (\frac{1}{2} - \frac{\varepsilon}{4})n + 6$ — as required by Lemma 6.3 — and V and $V^\parallel$ of σ_1 are both obtained by the the kind of procedure required in Lemma 6.4. It then follows by a union bound that if R is chosen by either σ_0 or σ_1 that, if A, B are subsets of $\{0,1\}^n$ of densities $\alpha, \beta \geq 2^{-(\frac{1}{2}-\varepsilon)n}$, then

$$\Pr_R\left[\frac{|A \times B \cap R|}{|R|} = (1 \pm 9 \cdot 2^{-\frac{\varepsilon}{4}n}) \cdot \alpha\beta\right] \geq 1 - 2 \cdot 2^{-\frac{\varepsilon}{4}n}.$$

Hence the same probability lower-bounds the event that $A \times B \cap R \neq \varnothing$. □

7 Partition number vs. communication complexity

In this section, we prove Theorem 1.4. This proof is similar to the proof of an analogous theorem in [GPW15]. Here we show a quadratic separation between logarithm of 1-partition number and deterministic communication complexity. The other separation, i.e., the separation between logarithm of partition number and deterministic communication complexity follows similarly.

Fix any function $s : \mathbb{Z} \to \mathbb{Z}$ and an $N \in \mathbb{Z}$, — we know from the assumption that $s(N) \leq \frac{\sqrt{N}}{\log n}$. We will exhibit a function F such that $F : \{0,1\}^N \times \{0,1\}^N \to \{0,1\}$, $\log \chi_1(F) = \tilde{O}(s(N))$ and $\mathcal{D}^{cc}(F) \geq s(N)^2$. Consider the function f for which [GPW15] proved a quadratic separation between unambiguous non-deterministic query complexity and deterministic communication complexity (*viz.* Theorem 5 of [GPW15]). f is a function which takes input from the alphabet $\Sigma = \{0,1\}^{k \times k} \times ([k] \times [k] \cup \{\perp\})$. i.e., the input size is $p \approx k^2 + 2\log k$. For this function f, which we refer to as GPW function, the authors show the following:

$$\mathcal{UN}^{dt}(\mathsf{GPW}) \leq 2k - 1,$$
$$\mathcal{D}^{dt}(\mathsf{GPW}) \geq k^2.$$

where $\mathcal{UN}^{dt}$ is the unambiguous non-deterministic query complexity. We can use our simulation theorem with Inner-product gadget (or Gap-Hamming gadget) (Theorem 1.3) to show the following immediately.

$$\mathcal{UN}^{cc}(\mathsf{GPW} \circ \mathsf{IP}) \leq (2k-1) \cdot n,$$
$$\mathcal{D}^{cc}(\mathsf{GPW} \circ \mathsf{IP}) \geq k^2 \cdot n.$$

Note that Theorem 1.3 holds even when $p = 2^{\frac{n}{200}}$. So, if we replace the value of n in the previous equation, we get,

$$\mathcal{UN}^{cc}(\mathsf{GPW} \circ \mathsf{IP}) \leq 200(2k-1) \cdot \log p \approx 200\sqrt{p - \log p}\log p \leq 200\sqrt{p}\log p,$$
$$\mathcal{D}^{cc}(\mathsf{GPW} \circ \mathsf{IP}) \geq 200k^2 \cdot \log p \approx 200(p - \log p)\log p \geq 100p\log p.$$

where the input size is $N' = 200p\log p$. We will set $N' = s(N)^2$ and will pad $(N - N')$ dummy bits to the function F to achieve our desired separation and input size. This, combined with the observation that $\log \chi_1(F) = \mathcal{UN}^{cc}(F)$, proves the theorem.

Acknowledgement

Part of the research for this work was done at the Institut Henri Poincaré, as part of the workshop *Nexus of Information and Computation Theories.*

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 616787. The first author is partially supported by a Ramanujan Fellowship of the DST, India and the last author is partially supported by a TCS fellowship.

The research leading to these results has received funding from the Foundation for Science and Tecnology (FCT), Portugal, grant number SFRH/BPD/116010/2016.

References

[AKK16] Andris Ambainis, Martins Kokainis, and Robin Kothari. Nearly optimal separations between communication (or query) complexity and partitions. In *Proceedings of the 31st CCC*, pages 4:1–4:14, 2016.

[BBCR13] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013.

[BBK$^+$13] Joshua Brody, Harry Buhrman, Michal Kouckỳ, Bruno Loff, Florian Speelman, and Nikolay Vereshchagin. Towards a reverse newman's theorem in interactive information complexity. In *Proceedings of the 28th CCC*, pages 24–33, 2013.

[BPSW05] Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. A direct sum theorem for corruption and the multiparty nof communication complexity of set disjointness. In *Proceedings of the 20th CCC*, pages 52–66, 2005.

[BR14] Mark Braverman and Anup Rao. Information equals amortized communication. *IEEE Transactions on Information Theory*, 60(10):6058–6069, 2014.

[BRWY13a] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct product via round-preserving compression. In *Proceedings of the 40th ICALP*, pages 232–243, 2013.

[BRWY13b] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. In *Proceedings of the 54th FOCS*, pages 746–755, 2013.

[CA08] Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. Technical Report TR08-002, Electronic Colloquium on Computational Complexity (ECCC), 2008.

[CDK$^+$17] Arkadev Chattopadhyay, Pavel Dvorák, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Lower bounds for elimination via weak regularity. In *34th Symposium on Theoretical Aspects of Computer Science, STACS 2017*, pages 21:1–21:14, 2017.

[Cha07] Arkadev Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *Proceedings of the 48th FOCS*, pages 449–458, 2007.

[Cha09] Arkadev Chattopadhyay. *Circuits, Communication and Polynomials*. PhD thesis, McGill University, 2009.

[CKLM17] Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Composition and simulation theorems via pseudo-random properties. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:14, 2017.

[dRNV16] Susanna F. de Rezende, Jakob Nordström, and Marc Vinyals. How limited interaction hinders real communication. In *Proceedings of the 56th FOCS*, 2016.

[Dru12] Andrew Drucker. Improved direct product theorems for randomized query complexity. *Computational Complexity*, 21(2):197–244, 2012.

[FF81] Peter Frankl and Zoltán Füredi. A short proof for a theorem of harper about hamming-spheres. *Discrete Mathematics*, 34(3):311–313, 1981.

[GLM$^+$15] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. In *Proceedings of the 47th STOC*, pages 257–266. ACM, 2015.

[GPW15] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *Proceedings of the 56th FOCS*, 2015.

[Har66] L.H. Harper. Optimal numberings and isoperimetric problems on graphs. *Journal of Combinatorial Theory*, 1(3):385 – 393, 1966.

[HJMR07] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. In *Proceedings of the 22nd CCC*, pages 10–23, 2007.

[Imp95] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proceedings of the 36th FOCS*, pages 538–545, 1995.

[Jai15] Rahul Jain. New strong direct product results in communication complexity. *Journal of the ACM*, 62(3):20, 2015.

[JKN08] Rahul Jain, Hartmut Klauck, and Ashwin Nayak. Direct product theorems for classical communication complexity via subdistribution bounds. In *Proceedings of the 40th STOC*, pages 599–608, 2008.

[JPY12] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for the two-party bounded-round public-coin communication complexity. In *Proceedings of the 53rd FOCS*, pages 167–176, 2012.

[JRS03] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *Proceedings of the 20th ICALP*, pages 300–315, 2003.

[JY12] Rahul Jain and Penghui Yao. A strong direct product theorem in terms of the smooth rectangle bound. Technical report, arXiv:1209.0263, 2012.

[KLL+15] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. *SIAM Journal on Computing*, 44(5):1550–1572, 2015.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[Lev87] Leonid A. Levin. One-way functions and pseudorandom generators. *Combinatorica*, 7(4):357–363, 1987.

[LSS08] Troy Lee, Adi Shraibman, and Robert Spalek. A direct product theorem for discrepancy. In *Proceedings of the 23rd CCC*, pages 71–80, 2008.

[LZ10] Troy Lee and Shengyu Zhang. Composition theorems in communication complexity. In *Proceedings of the 27th ICALP*, pages 475–489. Springer, 2010.

[Nor16] Jakob Nordström. Private communication, 2016.

[Pan12] Denis Pankratov. *Direct sum questions in classical communication complexity*. PhD thesis, Masters thesis, University of Chicago, 2012.

[RM99] Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. *Combinatorica*, 19(3):403–435, 1999.

[RY15] Anup Rao and Amir Yehudayoff. Simplified lower bounds on the multiparty communication complexity of disjointness. In *Proceedings of the 30th CCC*, pages 88–101, 2015.

[Sha03] Ronen Shaltiel. Towards proving strong direct product theorems. *Computational Complexity*, 12(1-2):1–22, 2003.

[She11] Alexander A Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 40(6):1969–2000, 2011.

[She12a] Alexander A. Sherstov. The multiparty communication complexity of set disjointness. In *Proceedings of the 44th STOC*, pages 525–548, 2012.

[She12b] Alexander A Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM Journal on Computing*, 41(5):1122–1165, 2012.

[She13] Alexander A. Sherstov. Communication lower bounds using directional derivatives. In *Proceedings of the 45th STOC*, pages 921–930, 2013.

[SZ09] Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5):444–460, 2009.

[VW08] Emanuele Viola and Avi Wigderson. Norms, xor lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4(1):137–168, 2008.

[WYY17] Xiaodi Wu, Penghui Yao, and Henry Yuen. Raz-mckenzie simulation with the inner product gadget. Technical Report TR17-010, Electronic Colloquium on Computational Complexity (ECCC), 2017.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the 11h STOC*, pages 209–213, 1979.

[Yao82] Andrew Chi-Chih Yao. Theory and applications of trapdoor functions (extended abstract). In *Proceedings of the 23rd FOCS*, pages 80–91, 1982.